International Journal of Engineering Science Invention
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org Volume 2 Issue 4 ǁ April. 2013 ǁ PP.14-15

Enterprise Data Mining: Issues & Solution

Majid Zaman

(Scientist, Directorate of Information Technology and Support System, University of Kashmir)

ABSTRACT :*Withallmostalltheenterprisegloballyhavinginstalleddatabaseapplicationsforautomationofactiviti es,databaseshavegrownoutinvolumeandhavemuchmoredatathenanticipatedbytheenterprises.Buttheprocessofretr ievingdesiredinformationisstillbeyondthecredibilityofcommonusersanddependencyisontheprogrammersandmana gersoftheapplicationsystemswhoarewellversedwithdatabasespecificquerylanguages.Theselanguagesscareawayn aiveendusers whoareleftatthemercyoftheprogrammers*
Itsnotonlyaboutdatabasesbutalsoaboutfilescollectedoveraperiodofyears.Enterprisegeneratesallsoughtoffilese.g.t xt,.pdf,.html,.jpgetcondaytodaybasis.Thefilesaresavedontheserversspreadacrosstheenterpriseandyetagaincommo nusersaredependentuponprogrammersandmanagersoftheapplicationsystems.Moreovertheenterpriseendusersare notatallawareoftheunderlyingarchitectureandneitherhavetheexpertiseinserveradministration.
Inthispaperiproposesolutionforenterprises,
whereinusercanretrievedesiredinformationwithouthavingtobotheraboutdatabaselanguagesand/orserver architecture.
Keywords–*Data Mining, key word, Search*

I. INTRODUCTION

Withthesuccessofsearchengineslikegoogle,yahooetcthekeywordbasedsearchhasbecomethebasisofinformationextr actionfromtheinternethowevernotmanysuchtoolshavebeendesignedforenterprises.Inkeywordsearchstringofwords normallyrefereedaskeywordsareprovidedbytheuserandbasedonthesekeywordssearchismadebythesearchenginesli kegoogleacrosstheglobalservers.Theuserispresentedwithlinksoftheserverswhereinkeywordshavematched.
Howeverinternetusersarenotgivenaccesstoenterprisedatabasesandneitherusersoftheinternetrequiresuchaccesstoda tabasesbutsamedoesnotholdtrueforenterpriseusers.Inenterprisemostofthedataisstoredinthedatabaseserversandcom monenterpriseusersisnothavingknowhowofdatabasespecificquerylanguagesandareleftatthemercyofdatabaseprogr ammers.itisneedofthehourtoprovidesimilarsearchparadigmforenterprisedatabasesuserswhocanquerybasedonkey wordsmuchthesamewayascommonmayqueryinternetongooglewithouthavinginformationnottheknowledgeofunde rlyingarchitecture[3].
Enterpriseusersshouldbeabletoquerydatabasewithouthavinganyknowledgeofdatabaseschemaneitherknowhowofu nderlyingquerylanguages.Dataisstoredindifferentdatabaseformatswheredatastorageandretrievaltechniquesarediff erentandnouniversalgenericrulesareapplicableeitheronstorageorretrieval,sothetooldevelopedshouldbesuchthatitd oesnotrelyuponuserinformationorability.
Dataisnotonlystoredindatabasesbutalsoindifferentfileformatse.g.txt,.html,.xmletc.weredataretrievalrulesarealtoge therdifferentandvaryfromformattoformat.Retrievingdatafromfilesrequiresenterpriseuserstobeawareoftheunderlyi ngfileschemaandalsotheknowledgeoftheretrievalmethodused,specifictothefileformat[4].
Enterpriseuserisalsoobligatorytohaveinformationretentionknowledge,astowhereinformationofhis/herinterestissto red.Itbecomesverycomplexastherecanbemfilesandntables.Enterpriseuseristheneithersupposedtodependonprogra mmerormemorieswheredataofhisinterestisstored,thingsgetovercomplicatedbecauseofdatareplication-samedatacanbestoredinnfileandntableatthesametime.Userdoesnotwanttodependonprogrammerneitherhavepatienc etomemorieswheredataisstored,userexpectstogivequerywithoutspecifyingwheredataisstoredandinwhichformatitis stored[5]

II. PROBLEM STATEMENT

Withtheadventofthetechnology, enterpriseacrosstheglobewhereinrushtocomputerizetheirprocess, howeverwithouthavingeyeonthefuturethesolutionswheredevelopedonheterogeneoussourcese.g

1. **EnterpriseBudget**developedonopen-sourcetechnologieslikelinuxapachemysqlphp
2. **EnterpriseHumanResource**developedonMSSQLmicrosoft.Nettecnolgies
3. **EnterpriseFinance**developedonOracle&java

Intheabovescenarioitsveryclearthatenterpriseishaving3variantdatabasesystemswhichinthemselveshasdifferenceu nderlyingarchitecture.Technicallyuserinterestedininformationisrequiredtohaveknowledgeoffalltheunderlyingdata base&databaseprogramminglanguages,whichisbeyondthecommonuserscapability.Usersinthiscasearealsorequired

toknowwheretheirinformationofinterestisstoredinordertoretrieveitasinwhichdatabaseandmorespecificallyinwhich table [5].

III. INTRANET

"Isthegenerictermforacollectionofprivatecomputernetworkswithinanorganization.Anintranetusesnetworktechnologiesasatooltofacilitatecommunicationbetweenpeopleorworkgroupstoimprovethedatasharingcapabilityandoverallknowledgebaseofanorganization'semployees.
IntranetsutilizestandardnetworkhardwareandsoftwaretechnologieslikeEthernet,WiFi,TCP/IP,WebbrowsersandWebservers" [1].

IV. PROPOSED SOLUTION

Theproposedsolutionmakesuseoforganizationintranetandexistingdatabaseservers.Theproblemandapplicabilityofsolutionisinitselfimperativethatmostoftheenterprisewhichhavedatabaseserversinstalledwillhaveintranettointerconnectorganizationinformationsystem.Inearly1990senterpriseacrosstheglobewantedtoautomatemostoftheirsolutions withdatabaseinthebackend,resultinginenormouscollectionofdata.
Theproposedsolutioniswebbasedwhereinwebapplicationdevelopedinterconnectsalldatabaseserversbasedontheir

1. ipaddress
2. username
3. password

irrespectiveoftypeofdatabasemanagementsystem.

"Adatabasemanagementsystem(DBMS)isthesoftwarethatallowsacomputertoperformdatabasefunctionsofstoring,retrieving,adding,deletingandmodifyingdata.Relationaldatabasemanagementsystems(RDBMS)implementtherelationalmodeloftablesandrelationships"[2].
ThoughDBMSallowsusertostore,retrieve,delete,updateetcdatafromthetablesofthedatabasehoweverourproposedsolutiononlyallowdapplicationtoretrievedataandnotmodifyit.Thereasonforallowingsolutiontoonlyretrievedataandnottoupdateormodifyitisbecauseourproposedsolutionismeantforminginformationfortheusersandisnotmeanttoworkastraditionaltransactionserver/OLAP.
Theproposedsolutionismeantforenterpriseusersandwillbemanagedbytheapplicationadministrator.Theadministrationoftheapplicationiswerethevariousdatabaseserverswillbemanagedbytheapplication,ndatabaseserverscanbeaddedtotheapplicationofvaryingtypesi.eOracle,mysql,MSSQLetc.,
Theapplicationstorestheipaddress,username&passwordalongwithtypeofDatabaseServere.gOracle,mysql,MSSQLoftheDatabaseServersfordiggingofinformation.
Theuserviewisgeneric where in
useronlyenterskeywordswhichareconvertedintothenquerybytheapplicationwherenisthenumberofdatabaseserversaddedtotheapplication.
Queriesareexecutedonndatabaseserver,andresultifavailableispresentedashyperlinktotheuser,incasetheuserquerymatchesmservers,thenthenmhyperlinksarepresentedtouser.User can view the result from specific servers by clicking the hyperlinks.

V. CONCLUSION

Inearly1990enterpriseacrosstheglobewantedtoautomatetheapplicationswithoutworryingaboutthearchitecture.Mostoftheenterprisesgloballyendeduphavingmultipleheterogeneoussources,andcullingoutusefulinformationfortheusersoftheenterprisebecameachallengeinitself.Thesolutionpresentedinthispaperisdesignedtomeetuserrequirementswhereinusercanaccessdesiredinformationwithouthavingtobotheraboutunderlyingarchitectureand/orschema.Inconclusiontheenterprisearenowmanagingandstreamliningthedatabuttillthenwecannotmakeenterpriseuserswait.

REFERENCES

[1] http://compnetworking.about.com/cs/intranets/g/bldef_intranet.htm
[2] http://databases.about.com/od/administration/g/dbms.htm
[3] R. Ashok Kumar, Dr Y. Rama Devi, "Efficient Approaches for Record level Web Information Extraction Systems". Published in International Journal of Advanced Engineering & Application, pp 161-164, Jan 2011 .
[4] Md. Sumon Shahriar and Jixue Liu, "Constraint-Based Data Transformation for Integration: An Information System Approach", International Journal of Database Theory and Application Vol. 3, No. 1,pp 85-92, March, 2010.
[5] J. Huang and E. Efthimiadis, "Analyzing and evaluating query reformulation strategies in web search logs". In Proceedings of CIKM, pp 77-86, ACM, 2009
[6] Ramakrishna Srikant, Sugato Basu, Ni Wang, Daryl Pregibon, "User browsing models: relevance versus examination". In Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 223-232, 2010